Decentralized MMSE Attacks in Electricity Grids

Iñaki Esnaola, Samir M. Perlaza, H Vincent Poor, Oliver Kosut

DECENTRALIZED MMSE ATTACKS IN ELECTRICITY GRIDS

Iñaki Esnaola[1,3]*, Samir M. Perlaza*[2,3]*, H. Vincent Poor*[3]*, and Oliver Kosut*[4].

[1]Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, UK.
[2]Institut National de Recherche en Informatique et Automatique (INRIA), Lyon, France.
[3]Department of Electrical Engineering, Princeton University, Princeton, NJ, USA.
[4]School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ, USA.

ABSTRACT

Decentralized data-injection attack construction with minimum-mean-square-error state estimation is studied in a game-theoretic setting. Within this framework, the interaction between the network operator and the set of attackers, as well as the interactions among the attackers, are modeled by a game in normal form. A novel utility function that captures the trade-off between the maximum distortion that an attack can introduce and the probability of the attack being detected by the network operator is proposed. Under the assumption that the state variables can be modeled as a multivariate Gaussian random process, it is shown that the resulting game is a potential game. The cardinality of the corresponding set of Nash Equilibria (NEs) of the game is analyzed. It is shown that attackers can agree on a data-injection vector construction that achieves the best trade-off between distortion and detection probability by sharing only a limited number of bits offline. Interestingly, this vector construction is also shown to be an NE of the resulting game.

Index Terms— Data-injection attacks, state estimation, game theory, decentralized attacks.

1. INTRODUCTION

The introduction of advanced sensing and communication infrastructure in electricity grids enables the implementation of applications and services envisioned in the smart grid paradigm but it also opens the door to cyber-security threats [1]. In this paper, data-injection attacks [2] against electricity grids are studied in a decentralized setting. The attack construction is formulated within a Bayesian framework in which the statistical structure of the state variables is exploited. With growing data mining and analysis capabilities provided by modern computing, it is reasonable to assume that network operators can learn the statistical structure of the system and incorporate it into a model of the underlying stochastic process governing the network [3]. Following the approach of [4], [5] and [6], the state variables are modeled as a multivariate Gaussian process whose second order moments are available to the attacker and the operator. The rationale for this is to consider a worst-case scenario for the operator. The validity of the Gaussian assumption about the distribution of the state variables is corroborated with real data for the case of low voltage distribution systems in [7].

Given the complexity and extent of most electricity grids, it is plausible to think of scenarios in which several attackers intrude upon a network at different locations. In this scenario, in which multiple attackers are present and/or limited communication is available among different instantiations of the same attacker, raises the notion of distributed attacks. Distributed attack and detection strategies are investigated in [4] and [8]. The decentralized system with different actors operating over a large number of processes poses a suitable framework for the exploration of game theoretic techniques. A comprehensive account of smart grid services and applications that can be tackled with game theory is given in [9]. In [10], centralized data-injection attacks are studied in a game theoretic setting in which the operator performs least squares estimation. Attack constructions that aim to manipulate market prices are modeled as a zero-sum game in [11]. However, the case in which several attackers disrupt the state estimation process in an uncoordinated way is still not well understood. Furthermore, the impact of making the statistical structure of the state variables available to attackers in decentralized settings has not been studied either. These issues are considered in this paper.

The next section describes the system model, including the estimation and detection procedures. The decentralized case and the properties of the resulting game are analyzed in Section 3. The paper ends with with concluding remarks in Section 4.

The work of S. M. Perlaza was supported in part by the European Commission under Marie Sklodowska-Curie Individual Fellowship No. 659316 (CYBERNETS). The work of H. V. Poor was supported in part by the U.S. National Science Foundation under Grants CMMI-1435778 and ECCS-1549881.

2. SYSTEM MODEL

Let $\mathbf{x} \in \mathbb{R}^N$ be a vector containing the state variables of a power system with N buses [12]. Assuming linearized system dynamics with M measurements corrupted by additive white Gaussian noise, the measurement vector $\mathbf{y}_o \in \mathbb{R}^M$ is given by

$$\mathbf{y}_o = \mathbf{H}\mathbf{x} + \mathbf{z}, \tag{1}$$

where $\mathbf{H} \in \mathbb{R}^{M\times N}$ is the Jacobian of the linearized system dynamics around a given operating point and $\mathbf{z} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_M)$ is additive white noise with power spectral density σ^2. The data-injection attack $\mathbf{a}$ is an M-dimensional deterministic vector introduced by an external attacker with the aim of disrupting the state estimation procedure. The attacker interferes with the measurements and modifies the observation model to

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z} + \mathbf{a}, \tag{2}$$

where $\mathbf{y} \in \mathbb{R}^M$ are the measurements available to the network operator that have been corrupted by the data-injection attack.

2.1. MMSE State Estimation

The aim of the network operator is to obtain an estimate $\hat{\mathbf{x}}$ of the state vector $\mathbf{x}$ using the measurements $\mathbf{y}$. In practical state estimation settings, linear estimators are preferred due to their simplicity, and thus the estimation procedure reduces to $\hat{\mathbf{x}} = \mathbf{L}\mathbf{y}$, given a linear estimation matrix $\mathbf{L}$. In the case in which the operator knows the underlying random process governing the state of the network, a common performance criterion for the estimation is to minimize the mean square error (MSE). In this case, the network operator uses a linear estimation matrix $\mathbf{M}$ that is the unique solution to the following optimization problem:

$$\mathbf{M} \triangleq \arg\min_{\mathbf{L}\in\mathbb{R}^{M\times M}} \mathsf{E}\left[\frac{1}{N}\|\mathbf{x} - \mathbf{L}\mathbf{y}\|^2\right], \tag{3}$$

where the expectation is taken with respect to the distributions of $\mathbf{x}$ and $\mathbf{z}$. Under the assumption that the network state vector $\mathbf{x}$ follows an M-dimensional real Gaussian distribution with zero mean and covariance matrix $\mathbf{\Sigma_{xx}}$, the minimum MSE (MMSE) estimation matrix is

$$\mathbf{M} = \mathbf{\Sigma_{xx}}\mathbf{H}^\mathsf{T}(\mathbf{H}\mathbf{\Sigma_{xx}}\mathbf{H}^\mathsf{T} + \sigma^2\mathbf{I})^{-1}, \tag{4}$$

and the MMSE estimate of the state vector $\mathbf{x}$ is

$$\hat{\mathbf{x}}_{\mathsf{MMSE}} = \mathbf{M}\mathbf{y}. \tag{5}$$

The aim of an attacker is to choose a data-injection vector $\mathbf{a} \in \mathbb{R}^M$ in order to obstruct the ability of the network operator to estimate the variables without being detected. Note that the impact of the data-injection vector $\mathbf{a}$ on the estimate $\hat{\mathbf{x}}_{\mathsf{MMSE}}$ is quantified by the second term on the right-hand side of the following equality:

$$\hat{\mathbf{x}}_{\mathsf{MMSE}} = \mathbf{M}(\mathbf{H}\mathbf{x} + \mathbf{z}) + \mathbf{M}\mathbf{a}. \tag{6}$$

The term $\mathbf{M}\mathbf{a}$ is referred to as the *excess distortion* induced by the attack vector $\mathbf{a}$ and is denoted by

$$\mathbf{x}_a = \mathbf{M}\mathbf{a} = \mathbf{\Sigma_{xx}}\mathbf{H}^\mathsf{T}(\mathbf{H}\mathbf{\Sigma_{xx}}\mathbf{H}^\mathsf{T} + \sigma^2\mathbf{I})^{-1}\mathbf{a}. \tag{7}$$

2.2. Attack Detection

As a part of grid management, a network operator performs bad data detection to identify corrupted measurements. This operation can be cast as a hypothesis testing problem with hypotheses

$$\mathcal{H}_0: \quad \text{There is no attack} \tag{8}$$

$$\mathcal{H}_1: \quad \text{Measurements are compromised.} \tag{9}$$

Assuming the operator knows that $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma_{xx}})$ it can obtain the joint probability density function of the measurements, $\mathbf{y}$, and the state variables $\mathbf{x}$. From (2) and the random model adopted to describe the state variables, it follows that the observations $\mathbf{y}$ are realizations of an M-dimensional real Gaussian random variable with covariance matrix

$$\mathbf{\Sigma_{yy}} = \mathbf{H}\mathbf{\Sigma_{xx}}\mathbf{H}^\mathsf{T} + \sigma^2\mathbf{I}, \tag{10}$$

and mean $\mathbf{a}$ when there is an attack; or zero mean when there is no attack. The resulting hypothesis testing problem compares the following hypotheses:

$$\mathcal{H}_0: \quad \mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma_{yy}}) \tag{11}$$

$$\mathcal{H}_1: \quad \mathbf{y} \sim \mathcal{N}(\mathbf{a}, \mathbf{\Sigma_{yy}}). \tag{12}$$

A worst case scenario approach is assumed for the attackers, namely, the operator knows the attack vector, $\mathbf{a}$, used in the attack. However, the operator does not know a priori whether the grid is under attack or not, which accounts for the need for an attack detection strategy. That being the case, the optimal attack detection strategy for the operator is to perform a likelihood ratio test $L(\mathbf{y}, \mathbf{a})$ with respect to the measurement vector $\mathbf{y}$. Under the assumption that state variables follow a multivariate Gaussian distribution, the likelihood ratio is given by

$$L(\mathbf{y}, \mathbf{a}) = \frac{f_{\mathcal{N}(\mathbf{0},\mathbf{\Sigma_{yy}})}(\mathbf{y})}{f_{\mathcal{N}(\mathbf{a},\mathbf{\Sigma_{yy}})}(\mathbf{y})} = \exp\left(\frac{1}{2}\mathbf{a}^\mathsf{T}\mathbf{\Sigma_{yy}}^{-1}\mathbf{a} - \mathbf{a}^\mathsf{T}\mathbf{\Sigma_{yy}}^{-1}\mathbf{y}\right), \tag{13}$$

where $f_{\mathcal{N}(\boldsymbol{\mu},\mathbf{\Sigma_{yy}})}$ is the probability density function of a multivariate Gaussian random variable with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$. Therefore, either hypothesis is accepted by evaluating the inequalities

$$L(\mathbf{y}, \mathbf{a}) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \tau, \tag{14}$$

where $\tau \in [0, \infty)$ is tuned to set the trade-off between the probability of detection and the probability of false alarm. The average probability that the network operator is unable to detect the attack vector $\mathbf{a}$ is

$$\mathsf{P_{ND}}(\mathbf{a}) = \mathsf{E}\left(\mathbb{1}_{\{\mathsf{L}(\mathbf{y},\mathbf{a})>\tau\}}\right), \tag{15}$$

where the expectation is taken over the state variables $\mathbf{x}$ and the noise $\mathbf{z}$, and $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function. Note that under these assumptions, $\mathbf{y}$ is a Gaussian random vector with mean $\mathbf{a}$ and covariance matrix $\mathbf{\Sigma_{yy}}$. Thus, the probability $\mathsf{P_{ND}}(\mathbf{a})$ of a vector $\mathbf{a}$ being a successful attack, i.e., a non-detected attack is given by [13]

$$\mathsf{P_{ND}}(\mathbf{a}) = \frac{1}{2}\mathsf{erfc}\left(\frac{\frac{1}{2}\mathbf{a}^{\mathsf{T}}\mathbf{\Sigma}_{\mathbf{yy}}^{-1}\mathbf{a} + \log\tau}{\sqrt{2\mathbf{a}^{\mathsf{T}}\mathbf{\Sigma}_{\mathbf{yy}}^{-1}\mathbf{a}}}\right). \tag{16}$$

3. DECENTRALIZED ATTACK CONSTRUCTION

This section describes the decentralized construction of data-injection attacks when K attackers are present. Let $\mathcal{K} = \{1, \ldots, K\}$ be the set of attackers that can potentially perform a data-injection attack on the network. Let also $\mathcal{C}_i$ be the set of sensors that attacker i controls. Assume that $\mathcal{C}_1, \ldots, \mathcal{C}_K$ are proper sets and form a partition of the set $\mathcal{M}$ of all measurement sensors. The set $\mathcal{A}_k$ of data attack vectors $\mathbf{a}_k$ that can be injected into the network by attacker $k \in \mathcal{K}$ is of the form

$$\mathcal{A}_k = \{\mathbf{a}_k \in \mathbb{R}^M : (\mathbf{a}_k)_j = 0 \text{ for all } j \notin \mathcal{C}_k, \mathbf{a}_k^T\mathbf{a}_k \leq E_k\}. \tag{17}$$

The constant $E_k < \infty$ represents the energy budget of attacker k. Let the set of all possible sums of the elements of $\mathcal{A}_i$ and $\mathcal{A}_j$ be denoted by $\mathcal{A}_i \oplus \mathcal{A}_j$. That is, for all $\mathbf{a} \in \mathcal{A}_i \oplus \mathcal{A}_j$, there exists a pair of vectors $(\mathbf{a}_i, \mathbf{a}_j) \in \mathcal{A}_i \times \mathcal{A}_j$ such that $\mathbf{a} = \mathbf{a}_i + \mathbf{a}_j$. Using this notation, let the set of all possible data-injection attacks be denoted by

$$\mathcal{A} = \mathcal{A}_1 \oplus \mathcal{A}_2 \oplus \ldots \oplus \mathcal{A}_K, \tag{18}$$

and the set of complementary data-injection attacks with respect to attacker k be denoted by

$$\mathcal{A}_{-k} = \mathcal{A}_1 \oplus \ldots \oplus \mathcal{A}_{k-1} \oplus \mathcal{A}_{k+1} \oplus \ldots \oplus \mathcal{A}_K. \tag{19}$$

Given the individual data-injection vectors $\mathbf{a}_i \in \mathcal{A}_i$, with $i \in \{1, \ldots, K\}$, the global attack vector $\mathbf{a}$ is

$$\mathbf{a} = \sum_{i=1}^{K} \mathbf{a}_k \in \mathcal{A}. \tag{20}$$

3.1. Choice of Utility Function

The aim of attacker k is to corrupt the measurements obtained by the set of meters $\mathcal{C}_k$ by injecting an error vector $\mathbf{a}_k \in \mathcal{A}_k$ that maximizes the damage to the network, i.e., the excess distortion, while avoiding the detection of the global data-injection vector $\mathbf{a}$. Clearly, all attackers have the same interest but they control different sets of measurements, i.e., $\mathcal{C}_i \neq \mathcal{C}_k$, for any pair $(i, k) \in \mathcal{K}^2$. For modeling this behavior, attackers use the utility function $\phi : \mathbb{R}^M \to \mathbb{R}$, to determine whether a data-injection vector $\mathbf{a}_k \in \mathcal{A}_k$ is more beneficial than another $\mathbf{a}'_k \in \mathcal{A}_k$ given the complementary attack vector

$$\mathbf{a}_{-k} = \sum_{i \in \{1,\ldots,K\}\setminus\{k\}} \mathbf{a}_i \ \in \mathcal{A}_{-k} \tag{21}$$

adopted by all the other attackers. The utility function ϕ is chosen considering the fact that an attack is said to be successful if it induces a non-zero distortion and it is not detected. Alternatively, if the attack is detected no damage is introduced into the network as the operator discards the measurements and no estimation is performed. Hence, given a global attack $\mathbf{a}$, the distortion induced into the measurements is $\mathbb{1}_{\{L(\mathbf{Hx}+\mathbf{z}+\mathbf{a},\mathbf{a})>\tau\}}\mathbf{x}_a^{\mathsf{T}}\mathbf{x}_a$. However, attackers are not able to know the exact state of the network $\mathbf{x}$ and the realization of the noise $\mathbf{z}$ before launching the attack. Thus, it appears natural to exploit the knowledge of the first and second moments of both the state variables $\mathbf{x}$ and noise $\mathbf{z}$ and consider as a metric the expected distortion $\phi(\mathbf{a})$ that can be induced by the attack vector $\mathbf{a}$:

$$\begin{aligned} \phi(\mathbf{a}) &= \mathsf{E}\left[\left(\mathbb{1}_{\{\mathsf{L}(\mathbf{Hx}+\mathbf{z}+\mathbf{a},\mathbf{a})>\tau\}}\right)\mathbf{x}_{\mathsf{a}}^{\mathsf{T}}\mathbf{x}_{\mathsf{a}}\right], &(22)\\ &= \mathsf{P_{ND}}(\mathbf{a})\ \mathbf{a}^{\mathsf{T}}\mathbf{\Sigma}_{\mathbf{yy}}^{-1}\mathbf{H}\mathbf{\Sigma}_{\mathbf{xx}}^{2}\mathbf{H}^{\mathsf{T}}\mathbf{\Sigma}_{\mathbf{yy}}^{-1}\mathbf{a}, &(23)\end{aligned}$$

where the expectation is taken over the distribution of the state variables $\mathbf{x}$ and the noise $\mathbf{z}$. Note that under these assumptions of global knowledge, this model considers the worse case scenario for the network operator. More specifically, the result presented in this section corresponds to a conservative case in which the attackers inflict the most harm.

3.2. Game Formulation

The benefit $\phi(\mathbf{a})$ obtained by attacker k not only depends on its own data-injection vector $\mathbf{a}_k$, but also on the data-injection vectors $\mathbf{a}_{-k}$ of all the other attackers. This becomes clear from the construction of the global data-injection vector $\mathbf{a}$ in (20), the excess distortion $\mathbf{x}_a$ in (7) and the probability of not being detected $\mathsf{P_{ND}}(\mathbf{a})$ in (16). Therefore, the interaction of all attackers in the network can be described by a game in normal form

$$\mathcal{G} = \left(\mathcal{K}, \{\mathcal{A}_k\}_{k\in\mathcal{K}}, \phi\right). \tag{24}$$

Each attacker is a player in the game $\mathcal{G}$ and it is identified by an index from the set $\mathcal{K}$. The actions player k might adopt are data-injection vectors $\mathbf{a}_k$ in the set $\mathcal{A}_k$ in (17). Given a vector of data-injection attacks $\mathbf{a}_{-k}$, player k aims to adopt a data-injection vector $\mathbf{a}_k$ such that the expected excess distortion $\phi(\mathbf{a}_k + \mathbf{a}_{-k})$ is maximized. That is,

$$\mathbf{a}_k \in \mathrm{BR}_k\left(\mathbf{a}_{-k}\right), \tag{25}$$

where the correspondence $\mathrm{BR}_k : \mathcal{A}_{-k} \to 2^{\mathcal{A}_k}$ is the best response correspondence, i.e.,

$$\mathrm{BR}_k\left(\mathbf{a}_{-k}\right) = \arg\max_{\mathbf{a}_k\in\mathcal{A}_k} \phi\left(\mathbf{a}_k + \mathbf{a}_{-k}\right). \tag{26}$$

From this perspective, a game solution that is particularly relevant for this analysis is the Nash equilibrium (NE) [14].

Definition 1 (Nash Equilibrium) *The data-injection vector* $\mathbf{a}$ *is an NE of the game* $\mathcal{G}$ *if and only if it is a solution of the fix point equation*

$$\mathbf{a} = \mathrm{BR}\left(\mathbf{a}\right), \tag{27}$$

with $\mathrm{BR} : \mathcal{A} \to 2^{\mathcal{A}}$ *being the global best-response correspondence, i.e.,*

$$\mathrm{BR}\left(\mathbf{a}\right) = \mathrm{BR}_1\left(\mathbf{a}_{-1}\right) \oplus \ldots \oplus \mathrm{BR}_K\left(\mathbf{a}_{-K}\right). \tag{28}$$

Essentially, at an NE, attackers obtain the maximum benefit given the data-injection vector adopted by all the other attackers. This implies that an NE is an operating point at which attackers achieve the highest expected distortion induced over the measurements. More importantly, any unilateral deviation from an equilibrium data-injection vector $\mathbf{a}$ does not lead to an improvement of the average excess distortion. Note that this formulation does not say anything about the exact distortion induced by an attack but rather it characterizes the average distortion. This is because the attack is chosen under the uncertainty of the state vector $\mathbf{x}$ and the noise term $\mathbf{z}$.

The following proposition highlights an important property of the game $\mathcal{G}$ in (24).

Proposition 1 *The game* $\mathcal{G}$ *in* (24) *is a potential game.*

Proof: The proof follows from the observation that all the players have the same utility function ϕ [15]. Thus, the function ϕ is a potential of the game $\mathcal{G}$ in (24) and any maximum of the potential function is an NE of the game $\mathcal{G}$. ■

In general, potential games [15] possess numerous properties that are inherited by the game $\mathcal{G}$ in (24). These properties are detailed by the following propositions

Proposition 2 *The game* $\mathcal{G}$ *possesses at least one NE.*

Proof: Note that ϕ is continuous in $\mathcal{A}$ and $\mathcal{A}$ is a convex and closed set; therefore, there always exists a maximum of the potential function ϕ in $\mathcal{A}$. Finally from Lemma 4.3 in [15], it follows that such a maximum corresponds to an NE. ■

3.3. Achievability of an NE

The attackers are said to play a sequential best response dynamic (BRD) if the attackers can sequentially decide their own data-injection vector $\mathbf{a}_k$ from their sets of best responses following a round-robin (increasing) order. Denote by $\mathbf{a}_k^{(t)} \in \mathcal{A}$ the choice of attacker k during round $t \in \mathbb{N}$ and assume that attackers are able to observe all the other attackers' data-injection vectors. Under these assumptions, the BRD can be defined as follows.

Definition 2 (Best Response Dynamics) *The players of the game* $\mathcal{G}$ *are said to play best response dynamics if there exists a round-robin order of the elements of* $\mathcal{K}$ *in which at each round* $t \in \mathbb{N}$*, the following holds:*

$$\mathbf{a}_k^{(t)} \in \mathrm{BR}_k\left(\mathbf{a}_1^{(t)} + \ldots + \mathbf{a}_{k-1}^{(t)} + \mathbf{a}_{k+1}^{(t-1)} + \ldots + \mathbf{a}_K^{(t-1)}\right). \tag{29}$$

From the properties of potential games (Lemma 4.2 in [15]), the following proposition follows.

Lemma 1 (Achievability of NE attacks) *Any BRD in the game* $\mathcal{G}$ *converges to a data-injection attack vector that is an NE.*

The relevance of Lemma 1 is that it establishes that if attackers can communicate in at least a round-robin fashion, they are always able to attack the network with a data-injection vector that maximizes the average excess distortion.

3.4. Cardinality of the set of NEs

Let $\mathcal{A}_{\mathrm{NE}}$ be the set of all data-injection attacks that form NEs. The following theorem bounds the number of NEs in the game.

Theorem 1 *The cardinality of the set* $\mathcal{A}_{\mathrm{NE}}$ *of NEs of the game* $\mathcal{G}$ *satisfies*

$$2 \leqslant |\mathcal{A}_{\mathrm{NE}}| \leqslant C \cdot \mathrm{rank}(\mathbf{H}) \tag{30}$$

where $C < \infty$ *is a constant that depends on* τ.

Proof: The proof of Theorem 1 can be found in [16]. ■

Theorem 1 shows that the set of attackers only need to share at most $\log_2 \lceil C \cdot \mathrm{rank}(\mathbf{H}) \rceil$ bits for coordinating an attack. Note that this exchange of information needs to take place only once and can be done offline. Interestingly, the authors have not been able to obtain any example in which $|\mathcal{A}_{\mathrm{NE}}| > 2$. In view of the numerical evidence the following conjecture is postulated.

Conjecture 1 *The set of NEs of* $\mathcal{G}$ *satisfies* $|\mathcal{A}_{\mathrm{NE}}| = 2$.

Conjecture 1 suggests that exchanging one bit of information is enough for the attackers to agree on the construction of an attack vector that leads to an NE.

4. CONCLUSION

This paper has established that decentralized attack construction strategies are feasible in a setting in which multiple attackers have limited communication. The trade-off between MMSE estimation and attack detection has been used to propose a novel utility function. This utility function gives rise to a game theoretic formulation that models the interaction among multiple attackers in the system. We have shown that the resulting game is a potential game, proved the existence of at least two NEs, and shown that the number of NEs is upper bounded by a finite number that depends on the network characteristics. Therefore the attackers cannot agree on an attack construction that will lead to an NE without coordination. Nonetheless, exchanging a finite number of bits offline enables the attackers to agree on a strategy that leads to an NE.

5. REFERENCES

[1] E. Hossain, Z. Han, and H. V. Poor, *Smart Grid Communications and Networking*, Cambridge University Press, 2012.

[2] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proc. ACM Conference on Computer and Communications Security*, Chicago, IL, USA, Nov. 2009, pp. 21–32.

[3] M. Ozay, I. Esnaola, F. T. Yarman Vural, S. R. Kulkarni, and H. V. Poor, "Machine learning methods for attack detection in the smart grid," *IEEE Trans. Neural Netw. Learn. Syst.*, to appear.

[4] S. Cui, Z. Han, S. Kar, T. T. Kim, H. V. Poor, and A. Tajer, "Coordinated data-injection attack and detection in the smart grid: A detailed look at enriching detection solutions," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 106–115, Sep. 2012.

[5] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 645–658, Oct. 2011.

[6] Y. Huang, H. Li, K. Campbell, and Z. Han, "Defending false data injection attack on smart grid network using adaptive CUSUM test," in *Proc. Annual Conference on Information Sciences and Systems*, Princeton, NJ, USA, 2011, pp. 1–6.

[7] C. Genes, I. Esnaola, S. M. Perlaza, L. Ochoa, and D. Coca, "Recovering missing data via matrix completion in electricity distribution systems," in *Proc. IEEE International Workshop on Signal Processing Advances in Wireless Communications*, Edinburgh, UK, Jul. 2016.

[8] M. Ozay, I. Esnaola, F. T. Yarman Vural, S. R. Kulkarni, and H. V. Poor, "Sparse attack construction and state estimation in the smart grid: Centralized and distributed models," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1306–1318, Jul. 2013.

[9] W. Saad, Z. Han, H. V. Poor, and T. Basar, "Game-theoretic methods for the smart grid: An overview of microgrid systems, demand-side management, and smart grid communications," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 86–105, Sep. 2012.

[10] I. Esnaola, S. M. Perlaza, and H. V. Poor, "Equilibria in data injection attacks," in *Proc. IEEE Global Conference on Signal and Information Processing*, Atlanta, GA, USA, Dec. 2014, pp. 779–783.

[11] M. Esmalifalak, G. Shi, Z. Han, and L. Song, "Bad data injection attack and defense in electricity market using game theory study," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 160–169, Mar. 2013.

[12] A. Abur and A. Gomez Exposito, *Power System State Estimation: Theory and Implementation*, CRC Press, 2004.

[13] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York: Springer-Verlag, 1994.

[14] J. F. Nash, "Equilibrium points in n-person games," *Proc. National Academy of Sciences of the United States of America*, vol. 36, no. 1, pp. 48–49, Jan. 1950.

[15] D. Monderer and L. S. Shapley, "Potential games," *Games and Economic Behavior*, vol. 14, no. 1, pp. 124–143, May 1996.

[16] I. Esnaola, S. M. Perlaza, H. V. Poor, and O. Kosut, "Decentralized maximum distortion MMSE attacks in electricity grids," *INRIA, Lyon, Tech. Rep. 466*, Sep. 2015.